An Instrumental Variables Approach to Measuring the Effect of Body Weight on Employment Disability

John Cawley

Objective. To measure the effect of body weight on employment disability.
Data Sources. Female respondents to the National Longitudinal Survey of Youth (NLSY), a nationally representative sample of American youth, surveyed from 1979 to 1998, merged with data from the child sample of the NLSY.
Study Design. A series of probit models and probit models with instrumental variables is estimated with the goal of measuring the effect of body weight on employment disability. The two outcomes of interest are whether a woman reports that her health limits the *amount* of work that she can do for pay, and whether she reports that her health limits the *kind* of work that she can do for pay. The models control for factors that affect the probability of health limitations on employment, such as education, cognitive ability, income of other family members, and characteristics of children in the household. Self-reports of height and weight are corrected for reporting error.
Principal Findings. All else being equal, heavier women are more likely to report employment disability. However, this overall correlation may be due to any or all of the following factors: weight causing disability, disability causing weight gain, or unobserved factors causing both. Instrumental variables estimates provide no evidence that body weight affects the probability of either type of employment disability.
Conclusions. This study finds no evidence that body weight causes employment disability. Instead, the observed correlation between heaviness and disability may be due to disability causing weight gain or unobservable factors causing both disability and weight gain.

Key Words. Obesity, employment disability, instrumental variables

Members of the public health community have reacted to the rising prevalence of obesity by urging increased efforts at weight loss (e.g., Koplan and Dietz 1999). These arguments are motivated with reference to the adverse outcomes associated with high body weight. However, if the relationship between weight and these outcomes is not causal, even successful efforts at

weight loss may not decrease the adverse outcomes associated with obesity. To understand the potential benefits of weight loss, we must understand which adverse outcomes are caused by high weight.

One adverse outcome linked to obesity in previous research is employment disability. For example, Narbro, Jonsson, Larsson, et al. (1996) studied a sample of Swedish women and found that in the course of a year, obese workers were more than twice as likely to be on disability pension and had 1.5 to 1.9 times more work days lost to illness than non-obese workers.

Such studies typically assume that obesity causes disability. However, three possible explanations exist for the correlation between high body weight and employment disability. First, as is commonly assumed, high weight may cause disability. A second explanation is that employment disability causes high weight. This may be true if, for example, disability leads to inactivity and weight gain. A third explanation is that unobserved factors cause both weight gain and employment disability. For example, if someone has a tendency not to plan for the future, that person may invest little in his or her health, which may result in obesity and employment disability.

This study tests the first explanation, that high weight causes employment disability, using the method of instrumental variables (IV). Specifically, the weight of a child is used as an instrument for the weight of the child's mother, which exploits the genetic variation in weight.[1]

IV methods have been used to evaluate the consequences of a therapeutic intervention, such as the use of cardiac catheterization for patients suffering heart attacks (Newhouse and McClellan 1998; McClellan, McNeil, and Newhouse 1994). This study is concerned with evaluating the consequences of a different kind of treatment. Instead of a treatment given by a doctor to a patient, extra body weight represents a treatment that potential patients partly choose and partly have imposed upon them by genetics.[2]

The fact that body weight is partly a matter of choice makes assessment of the effect of body weight on outcomes difficult. The correlation observed in the population between body weight and an outcome reflects both the true

---

I thank Gary Becker, John Bound, Charlie Brown, Mike Chernew, Julie Cullen, Rachel Dunifon, David Grabowski, Will Manning, and David Meltzer, as well as two anonymous referees for their helpful and insightful comments. I gratefully acknowledge financial support from the Robert Wood Johnson Foundation.

Address correspondence to John Cawley, Ph.D., Robert Wood Johnson Scholar in Health Policy, 109 S. Observatory, SPH-2, M2240, University of Michigan, Ann Arbor, MI 48109-2029. This article, submitted to *Health Services Research* on January 4, 2000, was revised and accepted for publication on July 27, 2000.

effect of weight on the outcome and the special characteristics of the people who have made choices that led them to that body weight. The goal of this study is to estimate the causal effect of body weight on employment disability.

One way to estimate this effect is to conduct a randomized clinical trial: treat an experimental group with higher body weight and compare their disability experience with that of a control group. Such an experiment would be unethical, so I have sought a natural experiment. Specifically, I use the weight of a child as an instrument for the weight of the child's mother. My identifying assumption is that a child's weight is highly correlated with the weight of the child's mother through genetics, and that after all observable factors have been taken into account the child's weight has no independent effect on the probability that the child's mother experiences employment disability. This assumption is supported by a body of literature in behavioral genetics. My IV method exploits the fact that certain people are endowed with a propensity to be heavier than others. By examining how this unchosen variation in body weight correlates with employment disability, I will derive an estimate of the causal effect of weight on disability.

For the purposes of this article, employment disability is defined as health limitations on the amount or type of work one can do for pay. The employment disability associated with obesity is of considerable economic significance; Wolf and Colditz (1998) estimate that, in the United States in 1995, obesity-related illness resulted in the loss of $3.93 billion in productivity due to 39.2 million lost work days.

The question of whether body weight causes employment disability is likely to become more important with time. Recent studies have found dramatic increases in the prevalence of overweight and obesity in the United States during the 1980s[3] and 1990s.[4] The rise is expected to continue (Flegal et al. 1998).

DATA: NLSY

The National Longitudinal Survey of Youth (NLSY), designed to represent the entire population of American youth, consists of a randomly chosen sample of 6,111 U.S. civilian youths, a supplemental sample of 5,295 randomly chosen minority and economically disadvantaged civilian youths, and a sample of 1,280 youths on active duty in the military.[5] All youths were between 14 and 22 years of age when the first of annual interviews was conducted in 1979. Since 1994, interviews have been conducted every two years. Retention rates

for those NLSY respondents remaining eligible for interview have remained close to 90 percent during the 16 years of interviews.

The NLSY recorded the self-reported weight of respondents in 1981, 1982, 1985, 1986, 1988, 1989, 1990, 1992, 1993, 1994, 1996, and 1998. Reported height was recorded in 1981, 1982, and 1985. Given the age of the NLSY respondents in 1985 (21–28), height in that year is assumed to be the respondents' adult height.

These self-reports of weight and height include some degree of reporting error, which may bias coefficient estimates. Specifically, when only one regressor is measured with error, attenuation bias exists in the OLS estimate of the coefficient associated with that regressor. However, if multiple regressors are measured with error, there is no reliable rule about the sign of the bias in the coefficients of the variables measured with error (Judge, Griffiths, Hill, et al. 1985). I correct the NLSY measures of weight and height for reporting error using the method of Lee and Sepanski (1995) and Bound, Brown, and Mathiowetz (in press); see Appendix for details.[6]

This study uses two measures of body weight: (1) weight in pounds, controlling for height in inches; and (2) body mass index (BMI).[7] BMI, the standard measure of fatness and obesity in epidemiology and medicine, is defined as weight in kilograms divided by height in meters squared.[8] Goodness of fit measures of the models estimated in this study are similar for the two measures of weight.

Weight tends to rise with age, and there was a trend in the general population toward increased obesity during the 1980s. In order to distinguish the effects of weight from those of age and time, age, age squared, and indicator variables for time are included as regressors. Weight may also be affected by pregnancy. For this reason, women who are pregnant at the time they report their body weight are dropped from the sample.[9] The final sample consists of 28,345 observations on 2,937 women.

The instrument used in the IV section of this article is the BMI of a biological child aged six to nine.[10] A single observation of child BMI is used as an instrument for each observation of mothers' weight.[11] Although the instrument does not vary with time, the mother's weight does, so in each year a woman may have a unique value for instrumented BMI. The data on child weight and height come from the Child Supplement to the NLSY, which consists of all children born to NLSY female respondents living in their mother's household at the time of a child assessment interview who completed an interview. The NLSY Child Supplement was administered every two years between 1986 and 1998. All of the children in the NLSY child sample are

related biologically to their mothers, so they represent suitable instruments to gauge the genetic variation in BMI. The NLSY child sample does not include information on the children of male respondents to the NLSY, so the sample used in this article is limited to women.

The use of children to instrument for mothers' weight requires that the sample be limited to women who have borne children. The empirical results found for this sample may not generalize to all women, but 82.65 percent of all women in the NLSY had given birth by 1998, so the sample of women with at least one birth may not differ significantly from the entire population of women.

For the purposes of this article, all occupations are classified as either white collar or blue collar, using Census codes for occupation. White collar workers are those working in sectors described by the U.S. Census as Professional, Technical, or Kindred Workers; Non-Farm Managers and Administrators; Sales Workers; and Clerical and Unskilled Workers. The only unskilled workers in the last group are those in white-collar positions, such as cashiers, file clerks, bill collectors, and messengers.

This study investigates two outcomes of interest: whether the respondent is limited by her health in her type of work for pay, and whether she is limited by her health in her amount of work for pay. In each interview, NLSY respondents were asked: "(Are you/Would you be) limited in the KIND of work you (could) do on a job for pay because of your health?" and "(Are you/Would you be) limited in the AMOUNT of work you (could) do because of your health?" The words in parentheses reflect the fact that this question was asked both of working and nonworking women. The yes or no answers to these questions are the dependent variables in this study. For the sake of simplicity, a person who answers yes to one of the above questions will be described as disabled.

The following regressors are included in the disability regressions: age, age squared, number of children in the household, age of youngest child in the household, log of family income other than that earned by respondent, general intelligence,[12] and highest grade completed. Indicator variables included are for year, local unemployment rate, region of residence, white-collar job, married with spouse present, has been married and spouse not present, no child in household, enrolled in school, black, and Hispanic.[13] These regressors are included because they may, for health reasons, affect the probability that a woman limits the type or amount of her paid employment. I also include indicator variables for missing data associated with each regressor

Table 1: National Longitudinal Survey of Youth Females Summary Statistics

Variable	*Number of Observations*	*Mean*	*Standard Deviation*	*Minimum*	*Maximum*
Health limits kind of work	28345	.04	.2	0	1
Health limits amount of work	28345	.03	.17	0	1
Body mass index (corrected)	28345	25.41	5.86	7.71	65.66
Weight in pounds (corrected)	28345	148.26	35.49	48.82	415.82
Height in inches (corrected)	28345	64.03	2.35	50.76	73.31
Indicator: black	28345	.31	.46	0	1
Indicator: Hispanic	28345	.2	.4	0	1
General intelligence	27423	−.09	.95	−3.62	2.97
Highest grade completed	28213	12.3	2.12	0	20
Enrolled in school	28345	.11	.31	0	1
Age	28345	28.36	5.73	16	41
Indicator: local UE rate $< 6\%$	27618	.37	.48	0	1
Indicator: local UE rate $\geq 9\%$	27618	.24	.43	0	1
Indicator: northeast region	28126	.15	.35	0	1
Indicator: north central region	28126	.25	.43	0	1
Indicator: west region	28126	.19	.39	0	1
Year	28345	89.37	5.14	81	98
Indicator: married, spouse present	28342	.54	.5	0	1
Indicator: been married, but not M-SP	28342	.18	.39	0	1
Number of children in household	28344	.18	.38	0	1
Indicator: no children in household	28344	.18	.38	0	1
Age of youngest child	23013	4.6	3.99	0	23
Log of other family income	21081	9.45	1.61	−.18	13.92
BMI of selected child	28345	17.25	3.79	7.51	64.54
Age of selected child	28345	7.97	1.02	6	9
Indicator: selected child is male	28345	.5	.5	0	1

except the weight variable. Summary statistics of the variables used in this study are presented in Table 1.

NLSY sample weights are used in all estimations described in this article. The z statistics reported in the tables reflect that robust standard errors are calculated with clustering by individual to account for correlations in the error terms of each individual over time. This is a variation on the method of White (1980).

AN INDEX FUNCTION MODEL OF DISABILITY

The goal of this section is to generate a consistent estimate of the causal effect of weight on employment disability. I assume that an individual i is disabled

at time t if her health H_{it} falls below some critical limit H_i^*. Health is assumed to be a function of body weight W_{it} and other characteristics X_{it}. Specifically,

$$H_{it} = W_{it}\beta + X_{it}\gamma + u_{it}.$$

Health H is not observed, but individuals report whether they are disabled; I denote $D_{it} = 1$ if individual i is disabled at time t, and $D_{it} = 0$ otherwise. Formally, disability relates to underlying health in the following way:

$$D_{it} = 0 \text{ if } H_{it} \geq H_i^*$$

$$D_{it} = 1 \text{ if } H_{it} < H_i^*.$$

Normalizing H^* at $H = 0$, the probability that one is disabled is equal to

$$\begin{aligned}\Pr[D_{it} = 1|X_{it}] &= \Pr[H_{it} < 0] \\ &= \Pr[W_{it}\beta + X_{it}\gamma + u_{it} < 0] \\ &= \Pr[u_{it} < -W_{it}\beta - X_{it}\gamma].\end{aligned}$$

If one assumes that u is normally distributed, one can estimate the probability of disability as a function of weight W and characteristics X using probit regression.

Probit estimates of the effect of weight on the probability that health limits the *amount* of paid work that one can do are contained in the first two columns of Table 2. Only information associated with weight and height variables is reported; complete results for all regressors are available upon request from the author. Shown are the marginal probabilities and z statistics associated with the probit coefficients. The estimated coefficients on both BMI and weight in pounds are positive and statistically significant at the 10 percent level, indicating that heavier women are more likely to report that their health limits the kind of work that they can do for pay. If weight is normally distributed, these coefficients imply the following: if two women who differed only in that one was at the median and one was at the 95th percentile in weight in pounds, the heavier woman would have a 0.7 percent higher probability that health would limit the amount she could work for pay. If two otherwise identical women differed in that one was at the median and the other at the 95th percentile in BMI, the heavier woman would have a roughly 0.6 percent higher probability that health would limit her amount of work. To help put these numbers in perspective, other results of the model indicate that aging from 35 to 40 increases the probability that health limits the amount of work a woman can do by 0.6 percent.

Table 2: Probits and Probits with Instrumental Variables
Dependent Variable = 1 If Health Limits Amount of Work
NLSY Females
Marginal Probabilities and (Z Scores)

	Probit		*Probit IV*	
BMI	.0005		0	
	(1.91)		(−.01)	
Weight in lbs.		.0001		0
		(1.99)		(−.02)
Height in inches		−.0001		.0003
		(−.14)		(.29)
Number of observations	28345	28345	28345	28345
Log likelihood	−3587.74	−3587.14	−3587.51	−3586.89

Note: Other regressors include: general intelligence, highest grade completed, age, age squared, age of youngest child, log of family income other than the respondents' wages, number of children in the household, and indicator variables for local unemployment rate, currently enrolled in school, region, marital status, no children in household, year, and black and Hispanic. NLSY sample weights are used in all regressions. Robust standard errors are calculated with clustering by individual. Pregnant women are dropped from the sample. Probit results reported are the marginal probabilities associated with probit coefficients; z scores appear in parentheses. Probit with IV uses the method of Newey (1987). Probit IV standard errors are corrected according to Murphy and Topel (1985). Instruments are interactions of child BMI with child age and gender.

The first two columns of Table 3 contain probit estimates of the effect of weight on the probability that health limits the *kind* of paid work that one can do. For each measure of weight, the coefficient estimates are positive and statistically significant at the 5 percent level; again, heavier women are more likely to report health limitations on their employment. The magnitude of the marginal probabilities in Table 3 are roughly twice those in Table 2. Using the same example as earlier, a woman at the 95th percentile in weight in pounds, relative to an otherwise identical woman whose weight was at the median, would have a 1.4 percent higher probability of reporting that her health limits the kind of paid work that she can do. If BMI is the measure of weight, that figure is 1.0 percent. In comparison, aging from 35 to 40 increases the probability of such disability by 1.05 percent.

I next explain why IV estimation is necessary. Assume that weight W has the following reduced form:

$$W_{it} = Z_{it}\pi + X_{it}\delta + \varepsilon_{it},$$

where X is the same set of variables that appeared in the disability equation, Z is a set of variables correlated with weight but not the error term in the disability equation, and ε is the residual.

Table 3: Probits and Probits with Instrumental Variables
Dependent Variable = 1 If Health Limits Kind of Work
NLSY Females
Marginal Probabilities and (Z Scores)

	Probit		*Probit IV*	
BMI	.0009		−.001	
	(2.4)		(−.46)	
Weight in lbs.		.0002		−.0002
		(2.44)		(−.45)
Height in inches		−.0009		.0005
		(−.8)		(.23)
Number of observations	28345	28345	28345	28345
Log likelihood	−4805.86	−4805.11	−4803.71	−4802.93

Note: Other regressors include: general intelligence, highest grade completed, age, age squared, age of youngest child, log of family income other than the respondents' wages, number of children in the household, and indicator variables for local unemployment rate, currently enrolled in school, region, marital status, no children in household, year, and black and Hispanic. NLSY sample weights are used in all regressions. Robust standard errors are calculated with clustering by individual. Pregnant women are dropped from the sample. Probit results reported are the marginal probabilities associated with probit coefficients; z scores appear in parentheses. Probit with IV uses the method of Newey (1987). Probit IV standard errors are corrected according to Murphy and Topel (1985). Instruments are interactions of child BMI with child age and gender.

Weight W is likely to be correlated with the error term in the probit regression for disability; unobservable factors probably affect both weight and the likelihood of becoming disabled. One such unobservable factor could be the rate of time discount: people who place little value on the future may become obese because they prefer current consumption to dieting and may also become disabled because they take fewer safety precautions at work. If this is true, the error terms ε and u will be correlated.

Just as in a linear regression, a correlation between a regressor and the error term violates the assumptions behind the nonlinear regression model.[14] One can still generate a consistent estimate of the effect of body weight on disability if one identifies a set of variables Z that are correlated with body weight but not u, the error term in the disability equation. Given Z, one can calculate an IV estimate of the effect of weight on disability.

In this article, instrument Z is the weight of a woman's child. The identifying assumption is that the BMI of a child is correlated with W, the weight of the child's mother, and is not correlated with u, the residual probability that the child's mother suffers employment disability. I estimate a probit IV regression according to the method of Newey (1987). The standard

errors for probit models with IV have been corrected for estimation in the first stage; see Murphy and Topel (1985).

In the first stage of IV estimation, a measure of mother's weight (corrected for reporting error)[15] is regressed on eight interaction terms: the child's BMI times indicator variables for child gender and age. There are eight interaction terms because there are two sets, one each for male and female children, of four indicator variables for each age six through nine. I interact the child's BMI with the child's age and gender because I want to measure the extent to which the child is heavy for his or her age and gender. The regressors from the second stage of IV are also included in the first stage, where OLS is used to regress BMI and weight in pounds on the instruments.

A series of articles has been published outlining the harms of weak instruments. Bound, Jaeger, and Baker (1993, 1995) point out two problems. First, a weak correlation between the instrument and the endogenous variable will exacerbate any problems associated with a correlation between the instrument and the residual in the second stage regression. Second, the magnitude of finite sample bias in IV estimates approaches that of the OLS bias as the R^2 between the endogenous explanatory variable and the instruments approaches 0. They suggest that the R^2 and F-statistics from the first stage of IV be reported as approximate guides to the quality of the IV estimates. According to Staiger and Stock (1997), 10 is an acceptable value of the F-statistic associated with the hypothesis that the coefficients on the instruments in the first-stage regression of IV are jointly equal to 0.

The set of instruments used in this study meets the standard of Staiger and Stock. In the first stage of IV estimation, the hypothesis that the coefficients on the instruments are jointly equal to 0 is rejected. First-stage coefficients, F-statistics, and partial R^2 associated with the excluded instruments are presented in Table 4. When BMI is the endogenous regressor, the instruments have an F-statistic equal to 10.7 and $\Delta R^2 = .04$.[16] When the endogenous regressor is weight in pounds, the instruments have an F-statistic equal to 10.46 and $\Delta R^2 = .10$. In Table 4, all of the coefficients on the instruments are of the expected sign; a high BMI child (relative to other children of the same age and gender) is associated with a heavier mother (whether measured in BMI or pounds). Each of the eight interactions between child BMI, child age, and child gender is statistically significant at the one percent level.

However, there are additional requirements of an instrument. In particular, the instrument must not be correlated with the error term in the second stage of IV estimation; if it is correlated, the IV procedure has accomplished nothing, and may in fact have caused harm (Bound, Jaeger, and Baker 1993)

Table 4: First Stage of IV
NLSY Females
Coefficients and (T Statistics)

	Dependent BMI	*Variable Weight in Lbs.*
BMI of daughter aged 6	.39	2.31
	(8.02)	(7.99)
BMI of son aged 6	.4	2.34
	(7.68)	(7.5)
BMI of daughter aged 7	.36	2.13
	(7.91)	(7.92)
BMI of son aged 7	.35	2.05
	(8.43)	(8.34)
BMI of daughter aged 8	.35	2.1
	(8.47)	(8.5)
BMI of son aged 8	.33	1.96
	(8.41)	(8.42)
BMI of daughter aged 9	.34	1.97
	(8.75)	(8.62)
BMI of son aged 9	.35	2.05
	(8.74)	(8.67)
R^2	.17	.22
ΔR^2 of instruments	.04	.1
F-statistic of instruments	10.7	10.46
Number of observations	28345	28345

Note: Other regressors include: general intelligence, highest grade completed, age, age squared, age of youngest child, log of family income other than the respondents' wages, number of children in the household, and indicator variables for local unemployment rate, currently enrolled in school, region, marital status, no children in household, year, and black and Hispanic. NLSY sample weights are used. Robust standard errors are calculated with clustering by individual. Pregnant women are dropped from the sample. Instruments are interactions of child BMI with child age and gender.

because the instrumented variable is still endogenous. The identifying assumption of this article is that the BMI of a child is correlated with the weight of his or her mother and is uncorrelated with the residual in the mother's disability equation. The evidence in favor of this assumption is as follows: (1) No consistent pattern exists between *childhood* obesity and socioeconomic status; see the review in Sobal and Stunkard (1989). (2) There is no measurable effect of common household environment on body weight; see Stunkard, Sorensen, Hanis, et al. (1986); Price and Gottesman (1991); Sorensen et al. (1992); Sorensen and Stunkard (1993); Vogler, Sorensen, Stunkard, et al. (1995); and Maes, Neale, and Eaves (1997).[17] These studies indicate that all

of the similarity in weight between parents and children is genetic in origin. Environment *unshared* by family members affects weight and is the reason the OLS estimates are suspect and IV may be necessary, but environment *shared* by family members has no detectable influence on weight.

While it is impossible to confirm the null hypothesis that child BMI is uncorrelated with the residual in the mother's disability equation, examining whether instruments are correlated with observable factors believed to be correlated with the unobservable factors that affect the second-stage residual can be informative. To this end, the sample has been divided into two groups: those whose selected child has an above average BMI for the selected children of that age and gender, and those whose selected child has a below average BMI. Table 5 lists the means of the variables used in this study for these two groups.

The rows of Table 5 are divided into three sections; the first is devoted to measures of weight, the second to explanatory variables, and the third to outcome variables. Means of the two groups in BMI and weight in pounds appear in the first and second rows. Mothers with high BMI children are, on average, heavier by 2.3 units of BMI, or thirteen pounds. This is consistent with the assumption that the instruments are correlated with the endogenous variable.

The second section of the table presents means of the explanatory variables for the two groups. In general, the means of the two groups are quite similar. For example, the two groups differ in average general intelligence by .043 units (this measure has, by construction, a standard deviation of 1) and in average education by only a tenth of a year. In addition, log of family income other than that earned by the mother is 9.401 for women with relatively heavy children, and 9.481 for women with relatively light children.

The group means in Table 5 are unconditional. As a test of a conditional relationship between the instruments and key observed explanatory variables, mothers' general intelligence, education, and other family income was each regressed on the set of instruments and the other regressors in the disability regressions. (Tables of these results are available upon request from the author.) For none of the three variables was the set of instruments statistically significant at the 10 percent level, which is additional suggestive evidence consistent with the IV identifying assumption.

Table 5 indicates that black women are more greatly represented in the heavy child group (34.3 percent) than in the light child group (29.1 percent). This difference suggests that the model coefficients may vary across race, but tests indicate that the equality of coefficients across race cannot be rejected.

Table 5: NLSY Women
Means of Regressors by IV Group

Variable	*Mothers of Above-Average BMI Kids*	*Mothers of Below-Average BMI Kids*
Weight and Height Variables		
Body mass index (corrected)	26.803	24.535
Weight in pounds (corrected)	156.648	142.975
Height in inches (corrected)	64.093	63.987
Explanatory Variables		
General intelligence	−.12	−.077
Highest grade completed	12.234	12.339
Log of other family income	9.401	9.481
Enrolled in school	.106	.107
Age	28.339	28.381
Indicator: black	.343	.291
Indicator: Hispanic	.203	.194
Indicator: local UE rate $< 6\%$	.363	.381
Indicator: local UE rate $\geq 9\%$	.247	.241
Indicator: northeast region	.143	.149
Indicator: north central region	.234	.264
Indicator: west region	.186	.186
Year	89.34	89.382
Indicator: married, spouse present	.521	.545
Indicator: been married, but not M-SP	.175	.188
Number of children in household	1.679	1.763
Indicator: no children in household	.174	.18
Age of youngest child	4.654	4.56
Outcome Measures of Disability		
Health limits kind of work	.042	.041
Health limits amount of work	.03	.028
Number of observations	10965	17380

The other explanatory variables have means that are very similar across the two IV groups. The overall similarity in means of explanatory variables is consistent with the identifying assumption that the distribution of unobserved factors correlated with the residual in the mothers' disability regression are uncorrelated with the instruments.

The two disability outcome variables occupy the bottom section of Table 5. A comparison of mothers with high BMI children with mothers of low BMI children suggests that the mothers with a genetic predisposition to heaviness have only very slightly higher incidence of employment disability: specifically, 0.1 percent higher incidence of health limitations on kind of work

and 0.2 percent higher incidence of health limitations on amount of work. This comparison represents a preliminary and unconditional IV estimate of the effect of weight on disability.

In summary, the instruments used in this study appear to be strongly correlated with the endogenous regressor and not to be correlated with the residual in the woman's disability regression. The instruments have another positive quality. IV analysis only measures the effect of the endogenous regressor on the dependent variable for the population "treated" by the natural experiment. In many natural experiments, the treated population differs in important ways from the general population, and the IV estimate for the treated population may differ dramatically from the treatment effect on the entire population (Angrist, Imbens, and Rubin 1996). Using the BMI of a child as an instrument for the weight of the child's mother largely avoids this problem, because genetics affects the body weight of every person and over 80 percent of the women in the NLSY have had children.

The results of probit analysis with IV are contained in the third and fourth columns of Tables 2 and 3. The results are consistent across the two tables. While the probit coefficients were statistically significant and positive, the probit with IV coefficients are not statistically significant. In addition, the marginal probabilities associated with weight are essentially 0; in Table 2 they are 0 to the fourth decimal point, and they are small and negative in Table 3. The hypothesis that weight does not affect the probability of either type of employment disability cannot be rejected.

While the hypothesis that the IV coefficients on weight are equal to 0 cannot be rejected, the true coefficients may lie at the high end of the 95 percent confidence intervals. The marginal effect associated with the largest coefficient in the confidence interval represents an upper bound on the effect of weight on employment disability. These bounds indicate that it can be ruled out that an extra unit of BMI increases the probability of health limitations on amount of work by more than 0.25 percent or increases the probability of health limitations on kind of work by more than 0.31 percent. Analogously, it can be ruled out that an extra pound of weight increases the probability of health limitations on amount of work by more than 0.04 percent or increases the probability of health limitations on kind of work by more than 0.05 percent. Effects of weight on employment disability smaller than these bounds cannot be ruled out using the current sample. In order to further limit the confidence intervals around the weight coefficients, and therefore the range of possible effects, a larger sample or a more powerful instrument is required.

The conflict between the results generated by probit models and probit models with IV is resolved with a Hausman (1978) test for exogeneity, the results of which indicate that the hypothesis that mothers' weight is endogenous cannot be rejected.[18] This implies that the method of IV generates more consistent estimates, and should therefore be preferred.

SUMMARY

This study finds that, overall, heavier women face a higher probability that health limits the kind and amount of work they can do for pay. Probit regressions indicate that a woman at the 95th percentile in BMI, relative to an otherwise identical woman whose BMI was at the median, would have a 1.0 percent higher probability of health limitations on the kind of paid work she can do. By comparison, aging from 35 to 40 increases the probability of such disability by 1.05 percent. These correlations cannot be interpreted as causal, however, because disability may increase weight or unobserved factors may increase both weight and the risk of health limitations on employment.

The method of instrumental variables is used to estimate the causal effect of weight on employment disability. The IV results are such that the hypothesis that weight does *not* increase the probability of either type of employment disability cannot be rejected. Instead, the observed correlation between heaviness and disability may be due to disability causing weight gain or unobserved factors causing both.

These findings suggest that weight loss may not decrease the risk of employment disability. Moreover, the results imply that estimates of the cost of obesity should not include the costs associated with higher incidence of disability among the obese. This can have a large impact on the estimates; for example, Wolf and Colditz (1998) estimate that, in the United States in 1995, the total cost of obesity was $99.2 billion. Included in that total is $2.5 billion of lost work days by obese female workers. To the extent that these lost work days were not truly caused by, but rather are merely correlated with heaviness, Wolf and Colditz have overestimated the cost of obesity.

Calls for weight loss are often motivated by reference to the adverse outcomes associated with high body weight. However, if the relationship between weight and these outcomes is not causal, even successful efforts at weight loss may not decrease the adverse outcomes associated with obesity. Future research should seek to discover whether other costly conditions associated with obesity are actually caused by obesity.

APPENDIX: REPORTING ERROR IN WEIGHT AND HEIGHT

Weight and height are self-reported in the NLSY; reporting error in these variables has the potential to bias coefficient estimates. This appendix assesses the extent of reporting error in weight and height in the NLSY and corrects for it, using the Third National Health and Nutrition Examination Survey (NHANES III) (U. S. Department of Health and Human Services 1996). NHANES III, conducted in 1988–1994, was designed to obtain information on the health and nutritional status of the U. S. population through interviews and direct physical examinations. The NHANES III sample is a nationally representative sample of 33,994 persons aged two months and older; 31,311 of those respondents also underwent physical examinations. NHANES III is useful for the purposes of this study because it both asked respondents to report their weight and height, and, within four weeks, measured their weight and height. To assess the extent of reporting error in the NLSY, I examined the reported and actual weight and height of NHANES III respondents of the same age as the NLSY sample when they reported their weight and height (aged 17–40). There were 3,854 female NHANES III respondents in the NLSY age range.

In NHANES III, height and weight are reported such that NLSY-aged women tend to underreport their BMI by 1.5 percent. Underreporting of weight varies positively with actual weight; underweight women overreport their weight, whereas overweight women underreport it. No clear pattern of misreporting of height was found.

To correct for this reporting error, true height and weight in the NLSY are predicted using information on the relationship between true and reported values in the NHANES III. This strategy is outlined in Lee and Sepanski (1995) and Bound, Brown, and Mathiowetz (in press). If one has validation data, which in this case contains measures of true and reported weight and height (and, therefore, BMI), one can regress the true value of the variable on its reported value. The OLS coefficient on the reported value is then used in the primary data set; specifically, it is multiplied by the reported value to create an estimate of the true value. (This assumes "transportability," i.e., that the relationships between true and reported values are the same in both data sets.)

Measured weight was regressed on reported weight for a sample of NLSY-aged (i.e., 17–40) female respondents to NHANES III. This regression is estimated separately by race; actual weight is regressed on reported weight and its square (in deviations about race group–specific means); the intercept

is suppressed.[19] Reported weight and its square are strong predictors of actual weight; judging by the extremely high R^2 (each over .995), this model fits the data very well.

This process was repeated for height. Regressions of actual on reported height and its square (in deviations about race-specific means) were estimated separately by race.[20] Again, the extremely high R^2 (equal to 1 to the third decimal place) suggests that reported height and its square are outstanding predictors of actual height.

Self-reported height and weight in the NLSY are then multiplied by the coefficients on the reported values associated with the correct race-gender group in the NHANES III. The fitted values of BMI and weight in pounds, corrected for reporting error, are used throughout the article. All of the models in this article have also been estimated using reported BMI, with very similar results.

NOTES

1. This study focuses on women because the data include information about the children of women, but not men.
2. Heritability studies suggest that genetics accounts for as much as 70 percent of the variance in weight across people; see Yanovski and Yanovski (1999).
3. Flegal et al. (1998) find that the prevalence of obesity in the United States rose from 14.5 percent in 1976–1980 to 22.5 percent in 1988–1994. Their estimates are based on measured weight and height.
4. Mokdad, Serdula, Dietz, et al. (1999) find that the prevalence of obesity increased from 12.0 percent in 1991 to 17.9 percent in 1998. These estimates are based on self-reported height and weight.
5. Due to funding constraints, some members of the original sample are no longer being interviewed. After the 1984 surveys, interviewing ceased for 1,079 members of the military subsample; retained for continued interviewing were 201 respondents randomly selected from the entire military sample. Beginning with the 1991 survey, 1,643 economically disadvantaged white respondents from the supplemental sample are no longer being interviewed.
6. I have also estimated the models in this article without correcting for reporting error in height and weight and I find very similar results.
7. Other measures of body weight were considered, in particular, indicator variables for clinical weight classification (e.g., overweight and obese). However, the results of specification tests indicated that the hypothesis that outcomes were adequately explained by the linear measures of weight in pounds and BMI, as opposed to nonlinear measures of weight, could not be rejected. Another reason for not using indicator variables for weight status is that the Newey (1987) method of probit

with instrumental variables, which is used in this study, generates consistent estimates only for continuous endogenous regressors.

8. The U.S. National Institutes of Health classifies BMI as follows: below 18.5 is underweight, between 18.5 and 25 is healthy, between 25 and 30 is overweight, and over 30 is obese. See U.S. National Institutes of Health (1998).
9. Two questions in the NLSY are used to eliminate women who are pregnant at the time they report their weight. First, women were asked whether they were currently pregnant at the time of interview. Second, in some years they were also asked whether they had, in retrospect, been pregnant at the time of the last interview. Women who answered yes to either of these questions are dropped from the sample in the year of pregnancy.
10. Sorensen, Holst, and Stunkard (1992) find that the mother-child correlation in BMI has reached its adult level by the child's age seven.
11. The most recent measured height and weight for the eldest child between the ages of six and nine were the values chosen as instruments. If no measured values of weight and height were available in any year for any child, the mother's reports of child height and weight were used, if available. Of all children whose weights are used as instruments, 98.7 percent have measured values. In models run with an added indicator variable for mother report of weight and height, the coefficient on the indicator was not statistically significant.
12. General intelligence is a measure of cognitive ability derived from the ten Armed Services Vocational Aptitude Battery tests; see Jensen (1987) for a full description of this measure of cognitive ability.
13. At the 10 percent significance level, the hypothesis that the coefficients estimated in this study are equal across race cannot be rejected.
14. In nonlinear regression, if a regressor is correlated with the error term, the transformed regressor is also expected to be correlated with the error term; see Greene (1993).
15. IV estimation is often proposed as a method of generating consistent estimates of coefficients of variables measured with error. See, for example, Fuller (1987) or Greene (1993). However, such an approach requires one to find an instrument correlated with the true value of the variable measured with error and yet independent of the reporting error. Since, as shown in the Appendix, reporting error in BMI is a function of level in BMI, to assume that an instrument correlated with true BMI is uncorrelated with the reporting error in BMI is not reasonable. For this reason, I must still correct self-reported height and weight for reporting error before IV estimation.
16. Most estimates from U.S. data of the correlation between the adult BMI of a mother and the childhood or adolescent BMI of her child are in the range .21–.36. The correlation does not differ by the gender of the child. See Maes, Neale, and Eaves (1997), p. 334.
17. The article by Grilo and Pogue-Geile (1991), a comprehensive review of studies of the genetic and environmental influences on weight and obesity, concludes that "... only environmental experiences that are not shared among family members appear to be important. In contrast, experiences that are shared among

family members appear largely irrelevant in determining individual differences in weight and obesity." (p. 520).

18. Specifically, this test consisted of regressing one of the disability outcomes on mothers' fitted weight and mothers' residual weight. The coefficient on the residual weight is statistically significant at the one percent level, indicating that the hypothesis that weight is endogenous cannot be rejected. This was the finding for each of the two measures of disability and each of the two measures of mothers' weight.
19. The hypothesis that the coefficients are equal across race was rejected. The hypothesis that these coefficients are equal across age groups could not be rejected. The hypothesis that the coefficient on the squared term is equal to 0 was rejected, but the hypothesis that the coefficient on a cubic term is 0 could not be rejected.
20. The hypothesis that the coefficients are equal across race was rejected. The hypothesis that these coefficients are equal across age groups could not be rejected. The hypothesis that the coefficient on the squared term is equal to 0 was rejected, but the hypothesis that the coefficient on a cubic term is 0 could not be rejected.

REFERENCES

Angrist, J., G. Imbens, and D. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–55.

Bound, J., C. Brown, and N. Mathiowetz. In press. "Measurement Error in Survey Data." *Handbook of Econometrics,* Volume 4, edited by J. Heckman and E. Leamer. New York: Springer-Verlag.

Bound, J., D. A. Jaeger, and R. Baker. 1993. "The Cure Can Be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables." Technical Paper No. 137. Cambridge, MA: National Bureau of Economic Research.

———. 1995. "Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak." *Journal of the American Statistical Association* 90 (430): 443–50.

Flegal, K. M., M. D. Carroll, R. J. Kuczmarski, and C. L. Johnson. 1998. "Overweight and Obesity in the United States: Prevalence and Trends, 1960–1994." *International Journal of Obesity* 22 (1): 39–47.

Fuller, W. A. 1987. *Measurement Error Models.* New York: John Wiley and Sons.

Greene, W. H. 1993. *Econometric Analysis,* 2nd edition. New York: Macmillan.

Grilo, C. M., and M. F. Pogue-Geile. 1991. "The Nature of Environmental Influences on Weight and Obesity: A Behavioral Genetic Analysis." *Psychological Bulletin* 110 (3): 520–37.

Hausman, J. A. 1978. "Specification Tests in Econometrics." *Econometrica* 46 (6): 1251–71.

Jensen, A. R. 1987. "The *g* Behind Factor Analysis." *The Influence of Cognitive Psychology on Testing and Measurement,* edited by R. R. Ronning, J. A. Glover, J. C. Conoley, and J. C. Dewitt. Hillsdale, NJ: Lawrence Erlbaum Publishing.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T.-C. Lee. 1985. *The Theory and Practice of Econometrics,* 2nd edition. New York: John Wiley and Sons.

Koplan, J. P., and W. H. Dietz. 1999. "Caloric Imbalance and Public Policy." *Journal of the American Medical Association* 282 (16): 1579–81.

Lee, L., and J. H. Sepanski. 1995. "Estimation of Linear and Nonlinear Errors-in-Variables Models Using Validation Data." *Journal of the American Statistical Association* 90 (429): 130–40.

Maes, H. H. M., M. C. Neale, and L. J. Eaves. 1997. "Genetic and Environmental Factors in Relative Body Weight and Human Adiposity." *Behavior Genetics* 27 (4): 325–51.

McClellan, M., B. J. McNeil, and J. P. Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality?" *Journal of the American Medical Association* 272 (11): 859–66.

Mokdad, A. H., M. K. Serdula, W. H. Dietz, B. A. Bowman, J. S. Marks, and J. P. Koplan. 1999. "The Spread of the Obesity Epidemic in the United States, 1991–1998." *Journal of the American Medical Association* 282 (16): 1519–22.

Murphy, K. M., and R. H. Topel. 1985. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business and Economic Statistics* 3 (4): 370–79.

Narbro, K., E. Jonsson, B. Larsson, H. Waaler, H. Wedel, and L. Sjostrom. 1996. "Economic Consequences of Sick-Leave and Early Retirement in Obese Swedish Women." *International Journal of Obesity and Related Metabolic Disorders* 20 (10): 895–903

Newey, W. K. 1987. "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables." *Journal of Econometrics* 36 (3): 231–50.

Newhouse, J. P., and M. McClellan. 1998. "Econometrics in Outcomes Research: The Use of Instrumental Variables." *Annual Review of Public Health* 19: 17–34.

Price, R. A., and I. I. Gottesman. 1991. "Body Fat in Identical Twins Reared Apart: Roles for Genes and Environment." *Behavior Genetics* 21 (1): 1–7.

Sobal, J., and A. J. Stunkard. 1989. "Socioeconomic Status and Obesity: A Review of the Literature." *Psychological Bulletin* 105 (2): 260–75.

Sorensen, T. I. A., and A. J. Stunkard. 1993. "Does Obesity Run in Families Because of Genes? An Adoption Study Using Silhouettes as a Measure of Obesity." *Acta Psychiatrica Scandanavia* 370 (Supplement): 67–72.

Sorensen, T. I. A., C. Holst, and A. J. Stunkard. 1992. "Childhood Body Mass Index–Genetic and Familial Environmental Influences Assessed in a Longitudinal Adoption Study." *International Journal of Obesity* 16 (9): 705–14.

Sorensen, T. I. A., C. Holst, A. J. Stunkard, and L. Theil. 1992. "Correlations of Body Mass Index of Adult Adoptees and Their Biological and Adoptive Relatives." *International Journal of Obesity* 16 (3): 227–36.

Staiger, D., and J. H. Stock. 1997. "Instrumental Variables Regression With Weak Instruments." *Econometrica* 65 (3): 557–86.

Stunkard, A. J., T. I. A. Sorensen, C. Hanis, T. W. Teasdale, R. Chakraborty, W. J.

Schull, and F. Schulsinger. 1986. "An Adoption Study of Human Obesity." *New England Journal of Medicine* 314 (4): 193–98.

U.S. Department of Health and Human Services, National Center for Health Statistics. 1996. *National Health and Nutrition Examination Survey III, 1988–1994* [Computer file]. ICPSR version. Washington, DC: U.S. Department of Health and Human Services, National Center for Health Statistics [producer]; Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor].

U.S. National Institutes of Health, National Heart, Lung, and Blood Institute. 1998. *Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults.* Washington, DC: National Institutes of Health.

Vogler, G. P., T. I. A. Sorensen, A. J. Stunkard, M. R. Srinivasan, and D. C. Rao. 1995. "Influences of Genes and Shared Family Environment on Adult Body Mass Index Assessed in an Adoption Study by a Comprehensive Path Model." *International Journal of Obesity* 19 (1): 40–45.

White, H. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (4): 817–30.

Wolf, A. M., and G. A. Colditz. 1998. "Current Estimates of the Economic Cost of Obesity in the United States." *Obesity Research* 6 (2): 97–106.

Yanovski, J. A., and S. Z. Yanovski. 1999. "Recent Advances in Obesity Research." *Journal of the American Medical Association* 282 (16): 1504–06.